Vision Research 95 (2014) 1–10

Contents lists available at ScienceDirect

Vision Research

journal homepage: www.elsevier.com/locate/visres

Information extraction during simultaneous motion processing

Reuben Rideaux *, Mark Edwards

Research School of Psychology, The Australian National University, Canberra, ACT 0200, Australia

ARTICLE INFO

Article history:
Received 20 June 2013
Received in revised form 26 November 2013
Available online 10 December 2013

Keywords:
Multiple object tracking

ABSTRACT

When confronted with multiple moving objects the visual system can process them in two stages: an initial stage in which a limited number of signals are processed in parallel (i.e. simultaneously) followed by a sequential stage. We previously demonstrated that during the simultaneous stage, observers could discriminate between presentations containing up to 5 vs. 6 spatially localized motion signals (Edwards & Rideaux, 2013). Here we investigate what information is actually extracted during the simultaneous stage and whether the simultaneous limit varies with the detail of information extracted. This was achieved by measuring the ability of observers to extract varied information from low detail, i.e. the number of signals presented, to high detail, i.e. the actual directions present and the direction of a specific element, during the simultaneous stage. The results indicate that the resolution of simultaneous processing varies as a function of the information which is extracted, i.e. as the information extraction becomes more detailed, from the number of moving elements to the direction of a specific element, the capacity to process multiple signals is reduced. Thus, when assigning a capacity to simultaneous motion processing, this must be qualified by designating the degree of information extraction.

1. Introduction

Extensive research has been conducted on the perception of motion, most of which has concentrated on the processing of single motion signals (Nishida, 2011). However, outside the lab, multiple motion signals within the visual field are common, e.g. the cars and pedestrians at a busy traffic intersection. While it is clear that we are capable of processing these signals, the precise mechanism and capacity of this ability remains relatively unknown.

There are at least two stages in which multiple motion signals can be processed by the visual system: an initial stage in which signals are processed in parallel (simultaneous motion processing) followed by a sequential stage (Edwards & Greenwood, 2005; Mulligan, 1992). Using the above example of a busy traffic intersection, during the initial stage of multiple motion processing a number of people and/or cars could be processed simultaneously. Following this, attention could be shifted among any remaining moving objects to process them sequentially.

It is difficult to determine the degree to which simultaneous processing occurs while navigating through a busy traffic intersection, as the difference between processing distinct moving targets such as people or vehicles in series or in parallel can be hard to estimate. This is reflected in the current debate over the mechanism of multiple object tracking. While some theories suggest simultaneous processing occurs, e.g. the FINST model (Pylyshyn, 1989), others offer a sequential account (d'Avossa et al., 2006; Oksama & Hyönä, 2008). However, one clear example of where simultaneous motion processing occurs is during the perception of transparent motion. Transparent motion is defined as more than one velocity field in the same part of the visual space and is due to either partial occlusions of moving objects or overlapping semitransparent surfaces (Qian, Andersen, & Adelson, 1994), e.g. a school of fish swimming upstream through moving water. The ability to perceive both the movement of the fish in one direction and that of the water in the other is an example of simultaneous motion processing. Thus, it is not surprising that the first studies investigating simultaneous motion processing employed transparent motion stimuli to explore this phenomenon (Mulligan, 1992).

Using a modified global-motion stimulus (Newsome & Pare, 1988), Edwards and Greenwood (2005) demonstrated that the maximum number of transparent motion signals defined only by differences in direction which could be simultaneously processed was two. They proposed that this limit of two was due to the elevated signal intensity threshold, defined as the proportion of motion signals within a given area moving at one velocity, relative to all others (Edwards & Nishida, 1999; Snowden & Braddick, 1989). Whereas the threshold for detecting unidirectional motion is around 10–15%, transparent motion requires over 40% for each signal. They later confirmed this, showing that the initial limit of two could be extended to three by additionally defining the signals by differences in speed and depth (Greenwood & Edwards, 2006a, 2006b). In doing so they engaged speed and disparity tuned global motion pathways with independent pooling (Edwards, Badcock, &

* Corresponding author.
E-mail address: Reuben.Rideaux@anu.edu.au (R. Rideaux).

0042-6989/$ - see front matter Crown Copyright © 2013 Published by Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.visres.2013.11.007

Smith, 1998; Hibbard & Bradshaw, 1999; van Boxtel & Erkelens, 2006), allowing them to effectively double the available signal intensity.

A common characteristic among the aforementioned studies was the use of spatially spread-out/transparent motion stimuli. However, outside the lab, occurrences of encountering three or more of these kinds of motion signals simultaneously are extremely rare. In contrast, occurrences of three or more spatially localized motion signals within the visual field are relatively common, e.g. a busy traffic intersection. Thus, while a limit of three may exist for processing transparent motion signals, this may not extend to motion signals which are spatially localized. Indeed, we recently investigated this hypothesis by asking observers to differentiate between two temporal presentations of n and $n + 1$ spatially localized motion signals (Edwards & Rideaux, 2013). We found that observers were able to differentiate between presentations containing five and six motion signals, suggesting a capacity to simultaneously process five signals. Additionally, by either increasing or decreasing the signal intensity we were able to increase the capacity to six and reduce it to four respectively, demonstrating the important role that signal intensity continues to play in determining the limit of this process, even when the signals are localized. Although the results from the discrimination task suggest a simultaneous motion processing capacity of at least six, it remains unclear as to what information is actually extracted at this level.

Progressing from low to high detail information extraction, here we investigate observers' ability to identify: (a) the number of signals present; (b) the actual directions present; and (c) the direction of a specific element. By measuring the capacity to extract these different types of information from multiple motion signals, we aim to determine whether the resolution of processing during the simultaneous stage varies as a function of information detail.

The findings from this study will also have considerable impact on the current debate between simultaneous and rapid sequential processing models in the field of multiple object tracking. Research shows that about four spatially localized objects can be accurately tracked (Pylyshyn & Storm, 1988). By determining what information can be simultaneously processed and from how many signals, we can demonstrate the in/feasibility of a putative simultaneous tracking model.

2. Experiment 1: number of signals

We recently demonstrated that observers were capable of discriminating between presentations containing five and six motion signals (Edwards & Rideaux, 2013). While this indicates a capacity to simultaneously process at least five motion signals, it remains uncertain whether observers were aware of the actual number of signals present as opposed to simply being able to determine that one interval contained more signals than the other. The aim of this experiment was to determine the maximum number of signal directions observers are capable of identifying during the simultaneous stage.

2.1. Method

2.1.1. Observers

Three observers were used, one of the authors (RR) and two others who were naïve with respect to the aims of the study. All had normal or corrected to normal acuity.

2.1.2. Apparatus

Stimuli were presented on a Phillips Brilliance 202P4 cathode-ray-tube monitor which was driven by a Cambridge Research Systems VSG 2/5 graphics card in a host Pentium computer. The monitor had a spatial resolution of 1024 × 768 pixels and a frame rate of 100 Hz.

2.1.3. Stimuli and procedure

A modified version of the stimulus used in our previous study was employed (Edwards & Rideaux, 2013). A single interval five-alternate forced-choice procedure was used. Each presentation contained between three (12 dots) and seven (28 dots) signal groups. The signal groups were defined by four dots arranged into a square pattern. These were formed by randomly selecting the location of the first dot while ensuring that it could move over the three motion frames without moving beyond the spatial extent of the viewing aperture. The remaining three dots were offset horizontally and vertically by 0.34° to form a square pattern. The total number of dots was kept constant at 60 by the addition of noise groups. Thus, in the three signal condition there were 12 (48 dots) noise groups and in the seven signal condition there were eight (32 dots) noise groups. All dots started off in the same four dot square pattern. The squares composed of signal dots kept their shape as they moved, as each dot making up that square moved in the same direction on each motion-frame transition, while squares composed of noise dots fell apart as each dot moved in a different randomly selected direction across the motion sequence. Each motion sequence consisted of three image frames, with each frame being presented for 60 ms.The observer's task was to indicate how many signal groups were contained within each presentation (from 3 to 7). A typical motion sequence with a signal level of five is shown in Fig. 1. The directions that each signal group moved in were randomly chosen from eight directions: the four cardinal and four diagonal directions. While no two signal groups could move in the same direction, the direction of the noise dots, fixed across each motion sequence, was unconstrained. That is, each noise dot could move in any direction over the full 360°. Observers ran 10 blocks of trials, with breaks as needed, each consisting of 50 presentations. The signal number conditions were randomly interleaved throughout each trial.

To prevent observers using just the static image in the last motion frame in each sequence to perform the task, a mask frame was presented at the end of each motion sequence. The mask consisted of 300 randomly positioned dots and was presented for 240 ms. In our previous study we found this mask to be effective (Edwards & Rideaux, 2013). The background had a mean luminance of 62 cd/m^2, and the dots had a positive Weber contrast of 20% and were 0.25° in diameter. The dots were displaced by 0.32° on each frame transition resulting in a speed of 5.3°/s and were presented in circular aperture with a diameter of 20°. The observer sat 50 cm from the monitor, with their head supported on a chin rest.

3. Results and discussion

The results of the three observers are shown in Fig. 2. Performance, percentage of trials the observers got correct, is plotted against the number of signals present. Given a 5AFC was used, threshold performance was set at halfway between chance (20%) and 100%, i.e. 60%. The pattern of results was similar for all observers. Only presentations containing up to four signals were performed at or above 60% (i.e. the 60% level fell within or below the 95% confidence intervals around the observer's performance level) meaning that observers could accurately identify the presence of up to four motion signals.

Additionally, two of the observers performed significantly above chance at a signal level of five. However, this can likely be attributed to a response bias within the higher signal level conditions (5, 6, and 7), indicated by performance at a signal level of

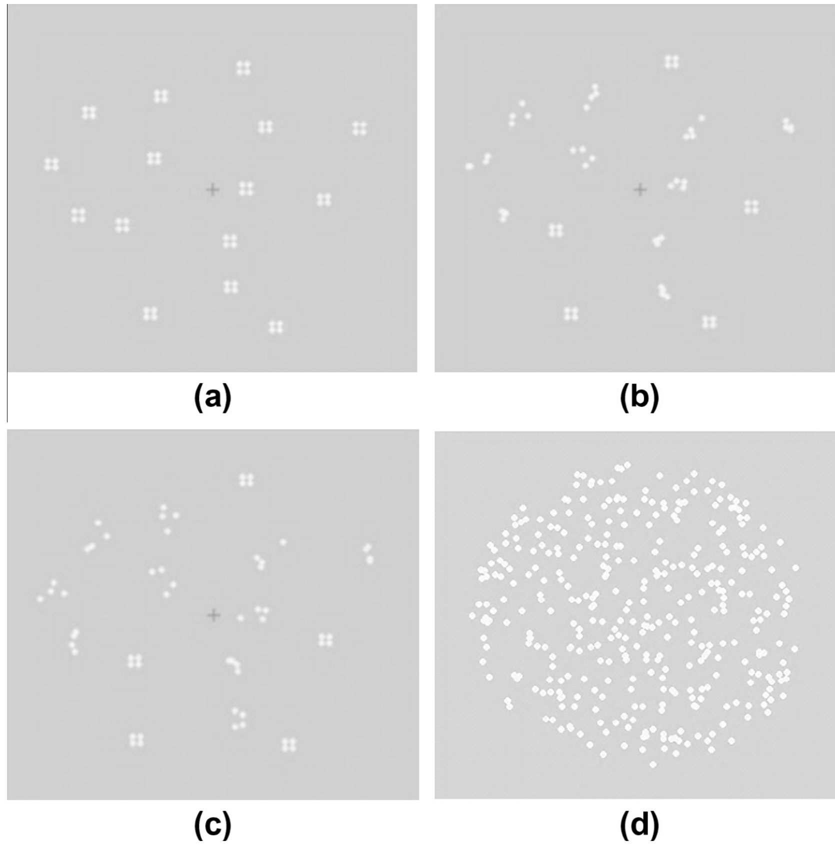

Fig. 1. An example of the stimuli used in Experiment 1. The images in the three-frame motion sequence are shown in (a) to (c) with a signal level of five. An example of the mask is shown in (d).

seven which has a similar magnitude of displacement from chance in the opposite direction, i.e. below chance.

These results are largely consistent with our previous findings, demonstrating multiple motion processing of more than three signals (Edwards & Rideaux, 2013). However, whereas in our previous study, which employed an n vs. $n + 1$ paradigm, we found a limit of five when we employed the same signal to noise levels as used here, the results from the current experiment suggest observers were only capable of identifying the number of signals present up to four. The difference suggests that the resolution of motion during simultaneous processing varies as a function of the type of information being extracted. When the task requires observers to discriminate between two presentations containing n vs. $n + 1$ numbers of motion signals, they can perform this accurately up to five vs. six signals. However when an observer is required to respond with the actual number of signals contained within a single presentation, they can only accurately perform this task with up to four signals present.

While the current experiment demonstrates that accurate numerosity judgements can be made with up to four signals, the next experiment determines to what extent motion information such as direction is extracted from these signals during simultaneous processing.

4. Experiment 2: directions present

Experiment 1 demonstrated that observers are capable of identifying the presence of up to four distinct motion signals from brief presentations. This shows that during simultaneous processing, information regarding the number of signals present within an area can be extracted. However, whether information regarding

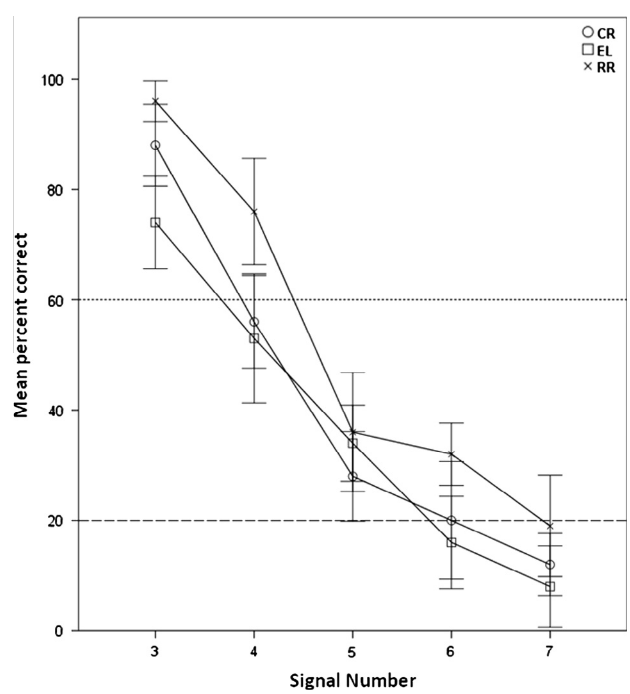

Fig. 2. Results for Experiment 1. The performance (percentage of responses that were correct) is plotted against the signal level. The dotted line indicates the above-chance performance threshold while the dashed line represents chance-level. Error bars indicate 95% confidence intervals.

the direction of motion is also extracted from these signals remains to be seen. The aim of Experiment 2 was to determine whether motion direction information is extracted during the simultaneous stage of multiple motion processing and if so, at what resolution this can be performed.

4.1. Method

4.1.1. Observers

Three observers were used, one of the authors (RR) and two others who were naïve with respect to the aims of the study. All had normal or corrected to normal spatial acuity.

4.1.2. Stimuli and procedure

The stimulus was the same as in Experiment 1 except that the noise dots and mask were removed and end frame altered. The end frame, the image presented after the motion sequence, was altered such that it now consisted of an arrow in the location of the fixation cross. The procedure was similar to that used in Experiment 1, except the observer's task now was to indicate whether the direction given by the arrow, out of eight possible directions, was present or absent within the preceding presentation. There was an equal chance of the target direction being present or absent. In the previous experiment the task required observers to identify the number of motion signals present. Noise dots were used to prevent observers from discerning the total number signal dots present without first recognizing them as the target elements, i.e. signal groups, through identifying their common motion. As the task in the current experiment required observers to identify the direction of the signal groups, discerning the total number of dots present would no longer act as a useful cue in performing the task. Thus, the noise dots were removed. However, by removing the noise dots the signal intensity across all signal levels was increased relative to the previous experiment and varied as a function of the number of signals present, i.e. the fewer signals the higher the signal intensity. As the static afterimage of the final frame could not be used as a cue to perform the task, the mask was also removed. Additionally, the signal level range was moved to two to six. An example of the motion sequence and end frame are shown in Fig. 3.

5. Results and discussion

The results of the three observers are shown in Fig. 4. The same performance criterion as used in Experiment 1 was employed, i.e. midway between chance and 100%. Given that a 2AFC was used, threshold performance was set at 75%. The pattern of results was similar for all observers. Only presentations containing up to three signals were performed at or above 75% (i.e. the 75% level fell within or below the 95% confidence intervals around the observer's performance level) meaning that observers could extract direction information from up to three signals.

At higher signal levels (4, 5 and 6) performance gradually declined but remained significantly above chance and at or above 75% for one of the observers (RR), the most experienced observer, at a signal levels four and five. Previous studies using transparent motion show that when the number of motion signals present exceeds the capacity of simultaneous motion processing, observers report perceiving noise and performance drops to chance (Edwards & Greenwood, 2005; Greenwood & Edwards, 2006a, 2006b). Here, the gradual decline in performance as a function of the number of signals present suggests that although the capacity of simultaneous processing was exceeded, observers extracted sufficient

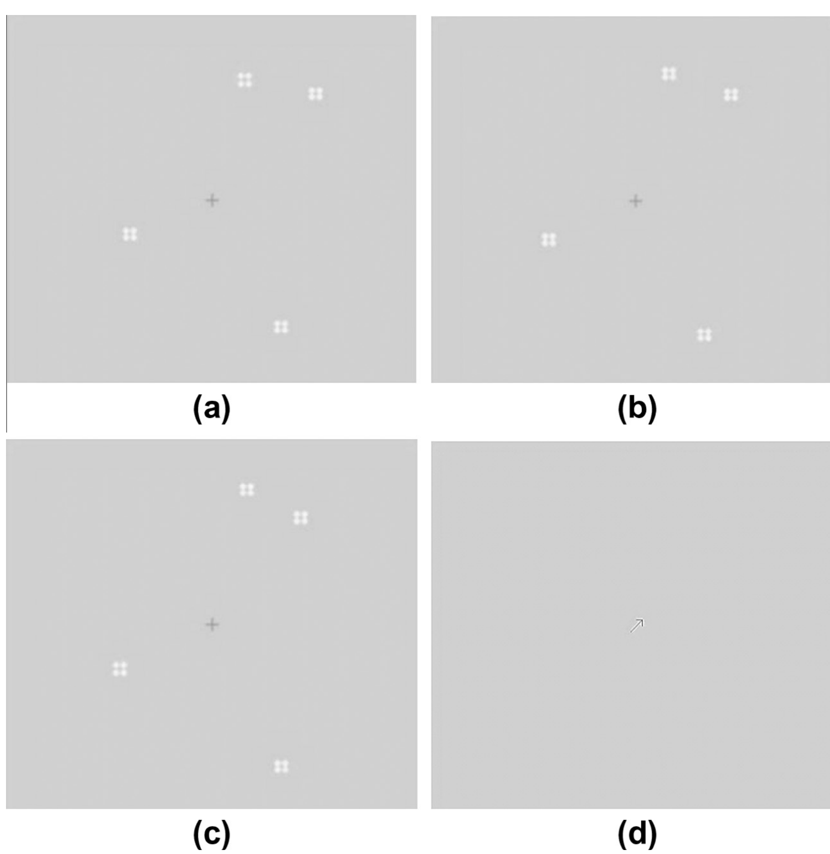

Fig. 3. An example stimuli used in Experiment 2. The images of the three frame motion sequence are shown in (a) to (c) with a signal level of four. An example of the post-cue target direction end frame is shown in (d).

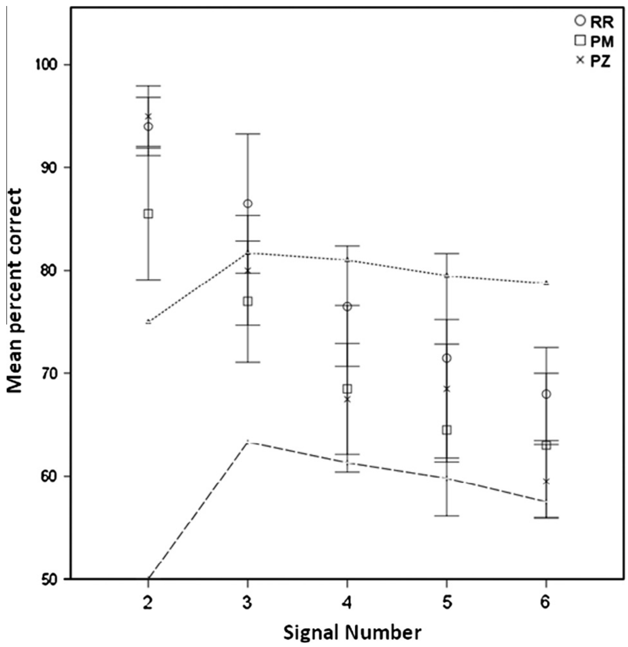

Fig. 4. Results for Experiment 2. Performance (percentage of responses that were correct) is plotted against the signal level. The dotted line indicates the adjusted chance-level while the dashed line indicates the adjusted above-chance performance threshold. Error bars indicate 95% confidence intervals.

direction information to perform at above chance-levels. These results are consistent with a two stage multiple motion processing mechanism where during the initial simultaneous stage information from a subset of signals is extracted leaving the remaining to be processed sequentially, i.e. as the number of signals increased the chance of the target direction being among the processed subset was reduced.

If observers were extracting information from a subset of signals when the number of elements exceeded the capacity to simultaneously process then modifying the chance and threshold levels at each signal level to reflect this interpretation may be more informative. For instance, with around 90% performance accuracy in the two signal level condition it is clear that observers are capable of processing at least two signals simultaneously. Thus, in the three signal level condition, if observers are processing only two signals, when the direction cue matched either of the two signals processed (on 33.3% of the trials, given the target was absent on half the trials), they would have a 90% chance of detecting this, i.e. 29.97% correct total responses. On the remaining 66.7% of trials the target direction would either be absent or present and matched to the unprocessed third signal; in which case observers would be expected to perform at chance levels, i.e. 33.3% of correct total responses. The sum of these two scores (63.3%) can now be applied as an adjusted chance level for the three signal level condition, with a corresponding adjusted threshold level of 81.7%. As performance in the three signal level condition exceeds this adjusted threshold, it is clear that observers were capable of simultaneously processing three signals. By applying the same principles to the remaining signal level conditions (4, 5 and 6), assuming that a subset of three signals is processed in each presentation, adjusted chance and threshold levels which may more accurately reflect the mechanism of simultaneous processing can be set. The addition of these adjusted chance/threshold levels, shown in Fig. 4, indicates that observers were capable of processing three signals, even when the number of signals present exceeded this.

Experiment 2 builds on the previous experiment by showing that during simultaneous motion processing observers are capable of extracting motion direction information from multiple signals. Furthermore, just as the limit found for numerosity identification of multiple signals (4) was lower than that of discrimination (5), the capacity found in the current experiment investigating motion direction extraction (3) is reduced further still. The reduced capacity is not due to a reduction in signal intensity as the total number of dots in the current experiment was less than that used in the previous. Thus, this finding provides additional credence to the notion that the resolution of simultaneous motion processing is also dependent upon the level of information extraction in question.

While Experiment 2 further demonstrates the degree of information extraction which occurs during simultaneous motion processing, i.e. the presence of a particular motion direction, it remains uncertain whether this information can then be bound to its corresponding signal. For example, following the multiple direction extraction seen in the current experiment, can the direction of a specific element be identified? Experiment 3 investigates direction binding.

6. Experiment 3: post-cue target location with iconic store inhibition

Experiment 3 aimed to determine whether during simultaneous motion processing, direction information which is extracted is bound to its corresponding signal.

6.1. Method

6.1.1. Observers

Three observers, including one of the authors (RR), were used. All had normal or corrected to normal vision.

6.1.2. Stimuli and procedure

The stimulus was the same as that used in Experiment 2, except a dynamic mask was introduced and the end frame altered. During pilot testing it was discovered that the task could be performed with relatively high accuracy even when the number of signals present far exceeded that which could be processed in the preceding experiments, i.e. 20. These results are characteristic of those found in partial report tasks, where a stimulus is briefly presented, followed by a cue indicating target items (Sperling, 1960). The observer's task during partial report tasks is to respond with information about the target items, i.e. orientation, colour, etc. Even though the number of elements present exceeds that which a person can perceive in such brief presentations, i.e. the span of apprehension (Cattell, 1885), if the proceeding cue is presented soon enough following this observers perform the task with high accuracy. This is due to iconic memory. Many types of information are stored in iconic memory; including orientation and spatial frequency (Magnussen, Idås, & Myhre, 1998), colour (Nilsson & Nelson, 1981), and motion direction (Demkiw & Michaels, 1976; Shooner et al., 2010; Treisman, Russell, & Green, 1975). Recovery of information from iconic memory demonstrates a relatively rich but transient capacity for storage and retrieval of briefly presented images (Sakitt, 1975, 1976). However, during partial report tasks a target subset of elements is selected, usually indicated by a cue following the presentation. Although all of the elements are stored in iconic memory, only information from the target subset is processed, i.e. encoded into working memory (Averbach & Coriell, 1961). The current experiment aimed to investigate the capacity of observers to extract information from multiple signals then indicate the direction of a single target element from within that group. The ability to use iconic memory to perform the task would allow observers to forgo processing multiple signals during the presentation; instead retroactively extracting the direction of the target signal once it has been cued. Thus, performance based on extraction of information from this store would not represent simultaneous motion processing as information is only extracted from one signal. In order to investigate the capacity to bind direction information which is extracted during simultaneous motion processing, the use of iconic memory must be prevented.

A static visual stimulus presented for duration of 130 ms or more persists in the iconic store for between 100 ms to 250 ms after its offset (Efron, 1970; Sperling, 1960). It has been suggested that the iconic representation of motion information may decay more slowly than static information and is therefore accessible for a longer duration (Demkiw & Michaels, 1976). Treisman, Russell, and Green (1975) found a significant reduction in partial report performance of motion between presentations where observers were cued at the stimulus offset and 1000 ms after the offset, however as no intermediate delays were tested the precise duration of persistence cannot be inferred. In contrast, Shooner et al. (2010) examined partial report performance at a range of delays between stimulus offset and 3000 ms and found a steep decline in performance between offset and 500 ms and similar performance at delays of 500, 1000, and 3000 ms. Thus, a 500 ms delay between the motion sequence offset and post-cue would be sufficient to prevent observers from using the location of the post-cue to extract the direction of the target signal from iconic memory. However, studies using static stimuli have shown that in the absence of a post-cue or mask, observers will process as much information as possible from the iconic store sequentially before it decays, referred to as "nonselective readout" (Averbach & Coriell, 1961). While this has only been examined using static images, it is possible that observers can also use nonselective readout to extract motion information. Thus, in order to prevent iconic memory being used either in conjunction with the post-cue or the potential to process signals sequentially, it must be interrupted at the offset of the stimulus using a mask.

Interruption of iconic memory occurs when the test stimulus is followed by a noise mask within the next 75 ms, depending on the conditions (Spencer, 1969; Spencer & Shuntich, 1970; Sperling, 1963). To achieve this, a dynamic mask was developed during pilot testing. Although there is considerable research on iconic memory and masks which are effective in inhibiting its use (for a comprehensive review see Breitmeyer & Ogmen, 2000 and/or Scheerer, 1973), there is little on iconic memory of motion and none regarding its masking. As a result, many masks had to be trialled during pilot testing before an effective one was found, i.e. the dynamic mask. The dynamic mask was created by drawing four static masks, as used in Experiment 1, and presenting each for 30 ms. This gave the impression of dots in random motion and looks similar to the black and white static observed on a television. As the aim of the current experiment was to determine the number of signal directions which can be simultaneously extracted during the brief presentation, the function of the dynamic mask was to interrupt iconic memory, not to disrupt the storage of information extracted during the presentation.

To demonstrate the effectiveness of the mask and that performance in this condition was a reflection of information extraction which occurred during simultaneous processing, one third of the presentations were followed by a 120 ms dynamic mask before the end frame was displayed, one third had a 500 ms delay during which only the fixation cross remained present and the remaining third had no delay. Performance in these conditions was expected to reflect simultaneous processing of the stimulus, a combination of simultaneous processing and sequential processing of signals stored in iconic memory and the use of the cue in conjunction with iconic memory, respectively. To observe the effect of signal level in the above three conditions, two signal level conditions, one with four signals present and one with eight, were randomly interleaved within each block. The end frame consisted of the final frame of the motion sequence with all but the target signal and fixation cross removed. Using arrow keys on the number pad of a keyboard, the observer's task was to indicate the direction of the post-cued target signal from eight possible responses: the 4 cardinal & 4 diagonal directions. Observers ran 10 trials, each of which consisted of 120 presentations. An example of the dynamic mask sequence and end-frame are shown in Fig. 5.

7. Results and discussion

The results of the three observers are shown in Table 1. The procedure employed an 8AFC design so chance performance was set at 12.5%. The pattern of results was similar for all observers. In all conditions observers performed significantly above chance. Given the interpretation taken from the previous experiment, that when presented with a number of signals exceeding the simultaneous processing limit the visual system will select a subset of these to process, expressing performance as the number of signals processed in each signal level condition is substantially more informative than what can be interpreted from a simple assessment of whether performance is above chance. However, in order to accurately translate performance into signals processed, accurate performance due to chance must first be removed. The higher performance is, the fewer correct responses are due to chance; between chance performance, where all are due to chance, and 100% performance, where none are due to chance. To determine the proportion of correct responses which are due to chance we must assume that incorrect responses represent responses which failed to be correct through chance. Given that chance was 12.5%, the proportion of incorrect responses represents the 87.5% of responses which failed to be correct through chance. Thus, the remainder of this 87.5% of responses are those that are correct due to chance. For example, if an observer's performance was 60%, the remaining 40% of (incorrect) responses represents 87.5% of responses which were guessed and failed to be correct due to chance. The proportion of the 60% correct responses which were correct due to chance can be determined by calculating the remainder of the of guessed responses, i.e. 40/87.5 * 12.5 = 5.7%. Once chance performance is removed, the adjusted performance can be used to accurately express the number of signals processed. The adjusted performance for each condition is shown in Table 1.

When performance is expressed as the number of signals processed from each presentation, by multiplying the number of signals presented by performance, the results show significantly better performance when more signals are presented in the no delay condition, (RR) $t(18) = 11.42$, $p < .001$, (PM) $t(18) = 5.61$, $p < .005$, and (CR) $t(18) = 10.62$, $p < .001$. In contrast, while performance was slightly better when fewer signals were presented in the delay and mask conditions, no significant differences were found.

Expressing performance as the number of signals processed is appropriate when the strategy of the visual system is to select a subset from those presented and accuracy reflects the probability of the target element being contained within this subset, i.e. if the visual system can process three signals simultaneously and is presented with four then with an increasing number of trials performance will approach 75%. However, if the strategy of the visual system is to use iconic memory to retroactively extract information from the target element, expressing performance as the number of signals processed is misleading as the number of signals which need to be processed to obtain high accuracy using this strategy is only one, i.e. the target. This is clearly the case in the no delay condition where performance expressed as the number of signals processed would suggest a higher capacity of signal processing

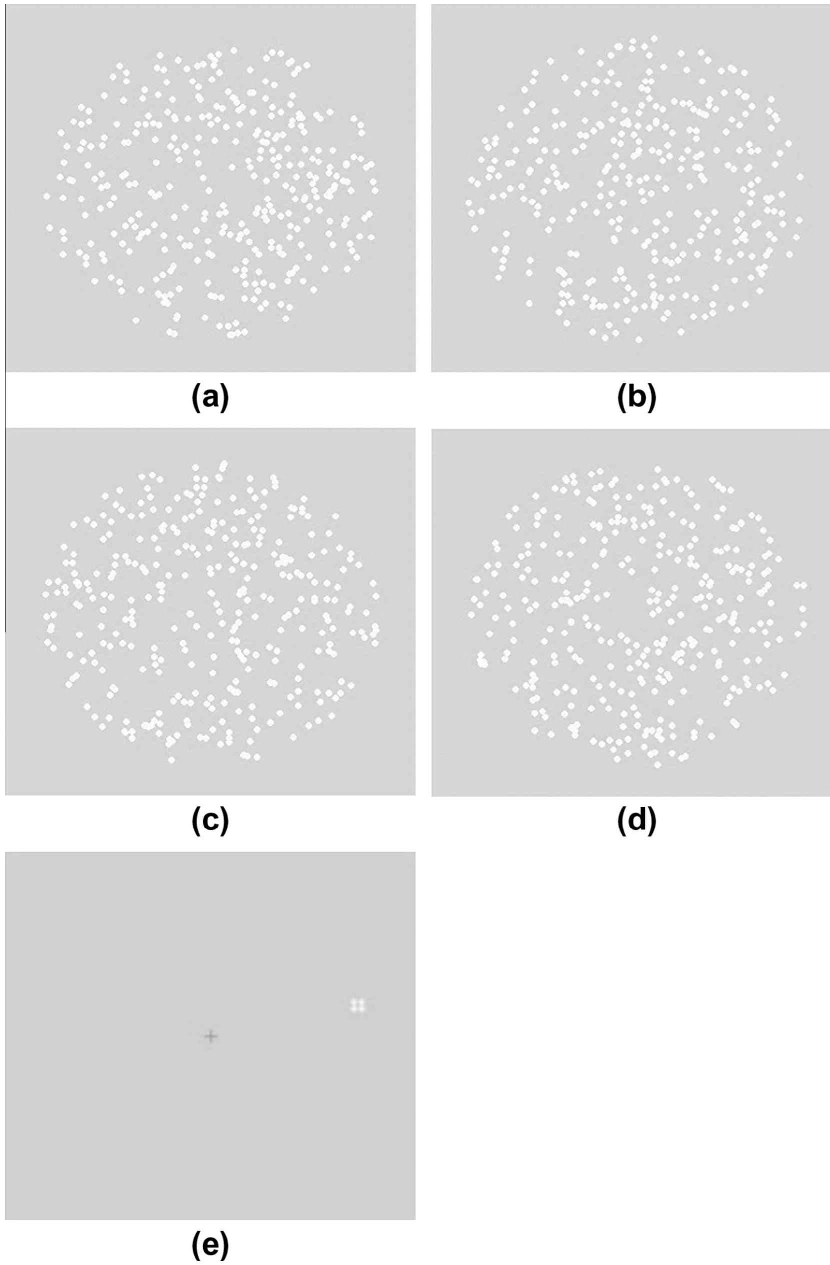

Fig. 5. An example of stimuli used in Experiment 3. The images of the four-frame dynamic mask are shown in (a) to (d) and a post-cue target location end frame is shown in (e).

when the signal level was increased. Given the strategy in the no delay condition employed iconic memory and thus accurate performance could be achieved from extracting information from only one signal, the results from this condition cannot be used to infer the capacity of simultaneous processing. In contrast, consistent performance between signal levels in the delay and mask conditions indicates that observers were capable of simultaneously extracting direction information, bound to a location, from a subset of up to three signals.

A defining characteristic of iconic memory is the relatively small effect which the number of items present has upon performance in a partial report task because only the target subset is processed (Sperling, 1960). However, given there were significant differences in performance between signal levels in the no delay condition for all observers, (RR) $t(18) = 6.21$, $p < .05$, (PM) $t(18) = 3.15$, $p < .05$, and (CR) $t(18) = 6.08$, $p < .05$, this indicates that increasing the number of signals presented had a negative effect on observers ability to extract motion information of a signal from iconic memory. This suggests that motion information stored in iconic memory may be less robust than static information.

Interestingly, performance represented as the number of signals processed was similar in the delay and mask conditions, suggesting two points. First, the additional duration in which it was proposed that signals stored in iconic memory may be sequentially processed in the delay condition did not significantly improve performance, (RR) $t(19) = -.4$, $p > .05$, (PM) $t(19) = .762$, $p > .05$, and (CR) $t(19) = 1.87$, $p > .05$. As research investigating nonselective readout has not yet been conducted using motion stimuli, this process may only apply to other characteristics of elements, e.g. orientation and colour. Secondly, the difference in signal intensity between the two signal levels did not have a significant effect on the number of signals which could be processed. Studies of simultaneous motion

Table 1
Observer performance in Experiment 3.

Observer	Condition	Signal level	Mean % correct	Adjusted mean	SD	Signals processed
RR	No delay	4	92	91	4.0	3.7
		8	77	74	7.7	5.9
	Delay	4	68	64	11.4	2.6
		8	46	38	11.0	3.1
	Mask	4	75	71	9.3	2.9
		8	42	33	12.1	2.7
PM	No delay	4	71	66	11.7	2.9
		8	54	47	9.3	3.8
	Delay	4	44	36	12.0	1.5
		8	27	17	14.9	1.4
	Mask	4	50	43	15.6	1.7
		8	30	20	9.9	1.7
CR	No delay	4	92	91	6.1	3.6
		8	75	72	9.9	5.8
	Delay	4	75	71	9.3	2.9
		8	43	35	14.3	2.8
	Mask	4	69	64	15.4	2.6
		8	40	32	15.5	2.6

processing using transparent motion stimuli have shown that the signal intensity required to process two signals is around three times greater than that needed to process one (Edwards & Greenwood, 2005). Similarly, we previously found that by reducing the intensity of spatially localized signals from around 7% to 5%, the number of signals observers were capable of discriminating between fell from five to four (Edwards & Rideaux, 2013). In the current experiment the signal intensity was halved, from 25% to 12.5%, between signal level conditions, yet the same number of signals appeared to be processed. While this is a relatively large reduction in signal intensity, as the signal intensity in the eight signal level condition (12.5%) still exceeds those tested in Edwards and Rideaux (2013) discrimination experiment, this may suggest that this is still sufficiently high not have an impact on the number of signals which can be processed. This further demonstrates that the mechanism of processing spatially localized signals is far more robust to noise than that used to process spatially spread-out signals (transparent motion).

Due to the nature of the task, i.e. motion signals moving in discrete directions, the ability to examine the degree of error regarding direction judgements was limited. This in turn restricts the capacity to compare the results of the current study with those of previous studies which have investigated this, such as Shooner et al. (2010). While this is an important aspect of multiple motion processing, mapping out directional judgement errors was peripheral to the aim of the present study, i.e. to determine the capacity of simultaneous motion information extraction.

During the current experiment it became apparent that iconic memory of motion may have been used to perform the task in Experiment 2. Thus, a control experiment was run using a signal level of four to compare performance with and without the dynamic mask. The three observers from the previous experiment were used. The results are shown in Fig. 6. Given that no significant differences between performance in the 'no delay' and 'dynamic mask' conditions were found, this indicates that iconic memory was not used to perform the task. The inability to use iconic memory to perform the task is likely due to the type of post-cue employed.

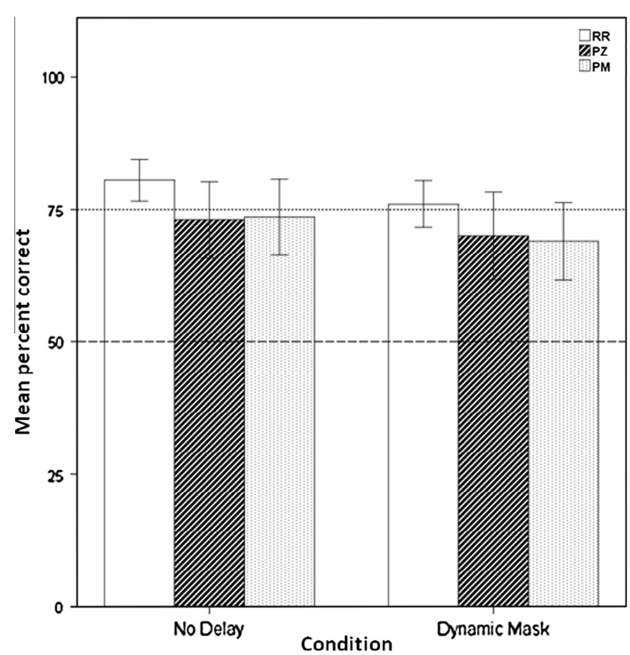

Fig. 6. Results for the control experiment. The performance (percentage of responses that were correct) is given for each condition; (a) no delay vs. dynamic mask and (b) absent vs. present. The dotted line indicates the above-chance performance threshold while the dashed line represents chance-level. Error bars indicate 95% confidence intervals.

8. General discussion

The main findings from the present study were that during the simultaneous stage of multiple motion processing, it is possible to extract the number of elements present, the actual directions of the signals, and the direction of a specific element. Furthermore, the resolution of simultaneous processing varies as a function of the information which is extracted. For instance, in our previous study we demonstrated that observers were capable of differentiating between presentations of up to 5 vs. 6 signals (Edwards & Rideaux, 2013), whereas here we show they were only capable of identifying the number of signals present up to four (Experiment 1). The resolution is further reduced to between two and three when observers are required to extract motion directions and identify the direction of a specific element (Experiments 2 and 3).

There are a number of implications from these results. The first is that while the simultaneous processing limit of three found by Greenwood and Edwards (2006) can be exceeded, the degree of information extraction suffers to the extent that motion direction information is lost. For instance, while the observer is aware of the presence of four distinct motion signals, they are only capable of extracting the direction of three during the simultaneous stage. Note that caution must be taken when comparing the limit found by Greenwood and Edwards (2006) to those in the current study. The task in their study required observers to discriminate between n vs. $n + 1$ transparent global motion signals. It remains unknown whether or not the same limit they found for a discrimination task (3) would apply to the extraction of global motion signal directions or if this would have a lower capacity, as demonstrated in the current study using localized motion signals. Further research is required to investigate the differences and similarities between the simultaneous processing of these two types of motion signals.

The second implication of the current study relates to the capacity to extract motion information from multiple signals even when the capacity is exceeded. As previously mentioned, studies in which transparent motion stimuli were used to measure the capacity of simultaneous motion processing found that when the capacity was exceeded, observers were unable to extract coherent motion and reported seeing noise (Edwards & Greenwood, 2005; Greenwood & Edwards, 2006a, 2006b). Thus, when presented with three transparent signals observers were able to extract the presence of all three, but when an additional signal was added they were unable to extract any. In contrast, when viewing a number of spatially distinct motion signals exceeding this same limit, observers were still capable of extracting direction information from a subset of these. This indicates that the visual system is capable of selecting a subset of motion signals, from a sample exceeding its capacity to simultaneously process, and extracting information from these before proceeding to sequentially process the remaining. While the mechanism of this process remains unknown, as this could not be achieved using transparent motion stimuli this suggests it operates on a spatially dependant basis, i.e. the area of motion extraction within the visual field is reduced to one which only contains up to the limit of signals which can be processed. However, it is likely that while in the current study this mechanism occurred passively, if the properties of the elements such as polarity were varied, i.e. half light/half dark, it is possible that an observer could actively process a subset using this characteristic as a cue (Edwards, 2009). We are currently investigating this possibility.

Finally, both of the abovementioned findings have important implications for multiple object tracking. Findings from multiple object tracking literature indicate that the maximum number of elements which can be tracked is around four, beyond which performance decreases (Allen et al., 2006; Pylyshyn & Storm, 1988). While the limit for this process has generally been established, the mechanism has not. While some theories suggest simultaneous processing occurs, e.g. the FINST model (Pylyshyn, 1989), others offer a sequential model (d'Avossa et al., 2006; Oksama & Hyönä, 2008). In the present study we demonstrate that observers are capable of simultaneously extracting motion information from up to three signals, providing support to a simultaneous processing model of multiple object tracking. It is important to note, however, that due to the structure of the signals used, i.e. four dots moving in the same direction, the task required the additional process of grouping. While this may explain the difference in between the capacity found in the present study (3) and those in multiple object tracking tasks (4) it also dictates that caution must be taken in comparing the two. However, stimuli used in multiple object tracking tasks incorporate both target and distractor signals. The distractor signals in these tasks are analogous to the remaining signals outside the subset selected for simultaneous processing during the task in the present study. Thus, while some differences exist between these tasks, i.e. grouping, the relevance of our findings to multiple object tracking is given extra credence by the demonstration that observers can simultaneously process a subset of signals in the presence of a larger sample.

Acknowledgment

This work was supported by Australian research Council Grant DP110104553.

References

Allen, R., Mcgeorge, P., Pearson, D. G., & Milne, A. (2006). Multiple-target tracking: A role for working memory? *The Quarterly Journal of Experimental Psychology, 59*(6), 1101–1116.

Averbach, E., & Coriell, A. S. (1961). Short-term memory in vision. *Bell System Technical Journal, 40*(1), 309–328.

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking: A comparison, review, and update. *Perception & Psychophysics, 62*(8), 1572–1595.

Cattell, J. M. (1885). On the time required for recognizing and naming letters and words, pictures and colors. *Philosophische Studien, 2*, 635–650.

d'Avossa, G., Shulman, G. L., Snyder, A. Z., & Corbetta, M. (2006). Attentional selection of moving objects by a serial process. *Vision Research, 46*, 3403–3412.

Demkiw, P., & Michaels, C. F. (1976). Motion information in iconic memory. *Acta Psychologica, 40*(4), 257–264.

Edwards, M. (2009). Common-fate motion processing: Interaction of the On and Off pathways. *Vision Research, 49*(4), 429–438.

Edwards, M., Badcock, D. R., & Smith, A. T. (1998). Independent speed-tuned global-motion systems. *Vision Research, 38*, 1573–1580.

Edwards, M., & Greenwood, J. A. (2005). The perception of motion transparency: A signal-to-noise limit. *Vision Research, 45*, 1877–1884.

Edwards, M., & Nishida, S. (1999). Global-motion detection with transparent-motion signals. *Vision Research, 39*, 2239–2249.

Edwards, M., & Rideaux, R. (2013). How many motion directions can be simultaneously perceived? *Vision Research, 76*, 11–16.

Efron, R. (1970). The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia, 8*(1), 37–55.

Greenwood, J. A., & Edwards, M. (2006a). An extension of the transparent-motion detection limit using speed-tuned global-motion systems. *Vision Research, 46*, 1440–1449.

Greenwood, J. A., & Edwards, M. (2006b). Pushing the limits of transparent-motion detection with binocular disparity. *Vision Research, 46*, 2615–2624.

Hibbard, P. B., & Bradshaw, M. F. (1999). Does binocular disparity facilitate the detection of transparent motion? *Perception, 28*, 183–191.

Magnussen, S., Idås, E., & Myhre, S. H. (1998).). Representation of orientation and spatial frequency in perception and memory: a choice reaction-time analysis. *Journal of Experimental Psychology. Human Perception and Performance, 24*(3), 707.

Mulligan, J. B. (1992). Motion transparency is restricted to two planes. *IVOS Supplement, 33*(4), 1049.

Newsome, W. T., & Pare, E. B. (1988). A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *The Journal of Neuroscience, 8*(6), 2201–2211.

Nilsson, T. H., & Nelson, T. M. (1981). Delayed monochromatic hue matches indicate characteristics of visual memory. *Journal of Experimental Psychology. Human Perception and Performance, 7*(1), 141.

Nishida, S. (2011). Advancement of motion psychophysics: Review 2001–2010. *Journal of Vision, 11*(5), 1–53.

Oksama, L., & Hyönä, J. (2008). Dynamic binding of identity and location information: A serial model of multiple identity tracking. *Cognitive Psychology, 56*, 237–283.

Pylyshyn, Z. (1989). The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition, 32*(1), 65–97.

Pylyshyn, Z. W., & Storm, R. W. (1988). Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial Vision, 3*, 179–197.

Qian, N., Andersen, R. A., & Adelson, E. H. (1994). Transparent motion perception as detection of unbalanced motion signals. I. Psychophysics. *The Journal of Neuroscience, 14*(12), 7357–7366.

Sakitt, B. (1975). Locus of short-term visual storage. *Science, 190*(4221), 1318–1319.

Sakitt, B. (1976). Iconic memory. *Psychological Review, 83*(4), 257–276.

Scheerer, E. (1973). Integration, interruption and processing rate in visual backward masking. *Psychologische Forschung, 36*(1), 71–93.

Shooner, C., Tripathy, S. P., Bedell, H. E., & Ogmen, H. (2010). High-capacity, transient retention of direction-of-motion information for multiple moving objects. *Journal of Vision, 10*(6), 1–20.

Snowden, R. J., & Braddick, O. J. (1989). The combination of motion signals over time. *Vision Research, 29*(11), 1621–1630.

Spencer, T. J. (1969). Some effects of different masking stimuli on iconic storage. *Journal of Experimental Psychology, 81*(1), 132.

Spencer, T. J., & Shuntich, R. (1970). Evidence for an interruption theory of backward masking. *Journal of Experimental Psychology, 85*(2), 198.
Sperling, G. (1960). The information available in brief visual presentations. *Psychological Monographs: General and Applied, 74*(11), 1–29.
Sperling, G. (1963). A model for visual memory tasks. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 5*(1), 19–31.
Treisman, A., Russell, R., & Green, J. (1975). Brief visual storage of shape and movement. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 699–721). London: Academic Press.
van Boxtel, J. J., & Erkelens, C. J. (2006). A single motion system suffices for global motion perception. *Vision Research, 46*, 4634–4645.